2020 Census
Program Management Review

Evaluations and Experiments-Related Projects

Kevin Deardorff, Program Manager

5.101 – Coding, Editing, and Imputation Study
8.102 – Administrative Records Fitness-For-Use
8.105 – Matching Process Improvement
8.107 – Administrative Records Modeling

March 8, 2013

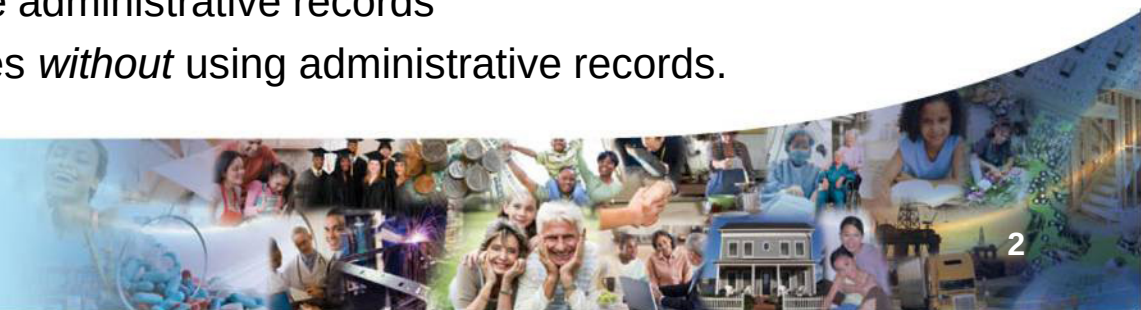
WBS 5.101 - Coding, Editing, and Imputation

Project Description:

Develop improved methods for missing data in the census:

- Improve methods for
  - count imputation: assign status of address (occupied, vacant, or delete?), and, if necessary, number of people in house
  - characteristic imputation (race, Hispanic origin, age, sex, relationship, tenure)
  - editing characteristic data
  - coding for write-ins of race and Hispanic origin

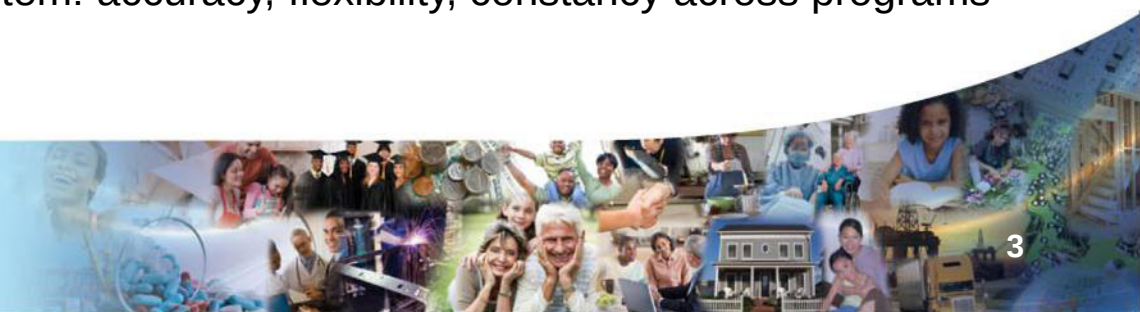
- Consider approaches that
  - include the use of available administrative records
  - enhance current procedures *without* using administrative records.



WBS 5.101 - Coding, Editing, and Imputation (Cont.)

Recent Accomplishments:

- Completed draft of study plan; presented to 2020 Leadership
- Discussed relationships and interdependencies with other teams
  - Presented plans, needs, and results
  - Worked with other teams to coordinate expectations with agreements and dates
- Reviewed schedule
  - Incorporated dependencies with other teams
  - Updated activities related to the 2013 Census Test and the 2014 Census Test
- Began discussion on metrics to evaluate procedures
  - reviewed metrics used in research on count imputation for 2010
  - for characteristic imputation and editing, discussed ways to evaluate important aspects of approach or system: accuracy, flexibility, constancy across programs



5.101 - Coding, Editing, and Imputation (Cont.)

Risk Highlights:

- If research assumptions differ from conditions in the 2020 Census, the results may not be valid.
- If we use administrative records for imputation, the availability of data for characteristics (e.g., race, Hispanic origin, tenure) will likely drop.
- If the team loses FTE resources through the normal attrition or transfers, work and results could be delayed.

5.101 - Coding, Editing, and Imputation (Cont.)

Challenges:

- The timeliness of policy decisions on research; we have begun discussions with the Policy and Communications Team.
- The ability to integrate software in real time (at the door; while responding via Internet) or near real time (over night).
- The ability to coordinate the various program requirements and software capabilities into one dependable, flexible system that runs as quickly as needed.

5.101 - Coding, Editing, and Imputation (Cont.)

Near-Term Project Focus Items:

- Continue to develop appropriate statistical metrics to evaluate results from alternatives; consult with other teams; focus on characteristic imputation and editing.
- Resolve questions relevant to our team about the upcoming tests.
- Continue developing new software; present updates within team.
- Incorporate administrative records into the research.
- Review results from research into administrative records modeling, as available; discuss implications on the extent and requirements of count and characteristic imputation.

WBS 8.102 - Administrative Records Fitness-For-Use

Project Description:

- Investigate administrative records (AR) sources to enhance data collection and processing methods for Nonresponse Followup (NRFU).
  - Explore agreement and disagreement in content across sources to develop approaches to resolve (such as modeling or business rules).
  - Produce recommendations for which AR sources are fit for operational uses such as count imputation, characteristic imputation, and modeling for NRFU.
- Project objectives

1) Describe NRFU addresses with AR available and those lacking AR.
2) Identify AR data that could support count and characteristic imputation.
3) Identify and address conflicting information across AR sources.
4) Assess quality of AR by source for intended uses (specifically NRFU).
5) Assess best time of year to acquire data relevant to Census Day and NRFU.

WBS 8.102 - Administrative Records Fitness-For-Use (Cont.)

Challenges Highlights:

- What is target level of geography and what metrics will gauge AR data quality?
- Can we build AR composites containing different sources for different geographies?
- Can we pursue state-specific (or region-specific or group-specific) data sources to enhance coverage of AR?
- Should we use un-validated person records (without anonymous Protected Identification Keys used for matching)? How would we do so?

WBS 8.102 - Administrative Records Fitness-For-Use (Cont.)

Other Challenges:

- Develop and implement strategies to help maintain staffing with the appropriate skill sets.
- Maintaining continued access to necessary and appropriate files.
- Develop and implement strategies to continue to follow protocols maintaining adherence to provider agreements as additional staff are granted access to the data.
- Continuance of our ability to solicit and acquire additional datasets to evaluate them & determine if their addition will better address our usage needs.

WBS 8.102 - Administrative Records Fitness-For-Use (Cont.)

Recent accomplishments:

- Initiated analysis of:
  - Description of NRFU addresses with AR at varying levels of geography.
  - Description of AR coverage of addresses responding by proxy.
  - Description of AR units that did not match to the Master Address File (MAF) or 2010 Census.
  - Use of Census Coverage Measurement (CCM) data to evaluate person records with conflicting address data.

- Designed a template for an Quality Criteria Checklist (QCC) to be used to describe the qualities and dimensions *of any administrative records or third party data* (ARTPD).

WBS 8.102 - Administrative Records Fitness-For-Use (Cont.)

Near term project focus items:

- Evaluation of using administrative records at the block level instead of the housing unit level.
- Design additional templates for Quality Criteria Checklists (QCC-2) to be used to describe the qualities and dimensions *of any composite of two or more administrative records &/or third party data* (ARTPD)
  - Description of which sources are best when combined and which sources could be excluded from models depending on intended use.
- Evaluation and improvement of "best address" selection when the same person is observed in multiple places.

WBS 8.105 – Matching Process Improvement

Project Description:

- Research and evaluate the methodology, techniques, and technology to improve matching of addresses and persons for the 2020 Census.
  - Research and assess matching techniques (including rule-based and probabilistic) to identify optimal methods by application.
  - Determine how to select cutoffs for probabilistic matching to optimize each application (i.e., how much agreement is needed between records to determine that they are actually a match).
  - Research error measurement in matching techniques, including determination of error and assumptions about causes.
  - Research and evaluate software to perform standardization and other processing to prepare data for matching.
  - Identify what data are needed to improve the quality of matching.

WBS 8.105 - Matching Process Improvement (Cont.)

Recent Accomplishments (1):

- Completed literature review and study plan.
- Updated schedule of key deliverables.
- Conducted brainstorming sessions to update the Project Risk Register and identified mitigation steps.
- Provided project action plans for the Knowledge Management Database to address recommendations from 2010 reports.
- Documented address standardization requirements and methodology from stakeholders throughout the Census Bureau.
- Identified test data requirements for comparing address standardization processes.

WBS 8.105 - Matching Process Improvement (Cont.)

Recent Accomplishments (2):

- Documented existing Census Bureau error measurement software.
- Held the inaugural meeting for the Comparing Matching Systems subteam.
- Obtained access to 2010 Census Coverage Measurement data and completed requests for access to administrative records data.

WBS 8.105 – Matching Process Improvement (Cont.)

Risk Highlights:

- If adequate and reliable computing resources with limited contention (servers, disk space) are not available, then research will be delayed or limited.
- If in-house or commercial software isn't compatible with other operating systems or other software used in a given decennial application, then implementation of recommendations may be limited.
- If we do not receive timely information about alternative designs and matching needs from other teams, then we can't fully anticipate matching needs for 2020 and matching research may be delayed or limited.
- If staff knowledgeable of clerical methods are not available for enough time, then improvements to automated methods based on lessons learned from clerical work will be limited and cost savings from reducing the clerical review workload will be limited.

WBS 8.105 - Matching Process Improvement (Cont.)

Challenges:

- If access to required administrative record files data is not received in full or in a timely manner, then matching research will be delayed or limited.

Near-term Project Focus (1):

- Compare address standardization options using evaluation criteria and test decks identified by the team.
- Identify refinements needed to current record linkage and error measurement methods.
- Continue evaluating matching cutoffs by reviewing 2010 Census Coverage Measurement clerical results and researching methods to automate the determination of matching cutoffs.
- Continue researching refinements to current methods used to identify person duplication in the census (e.g., modifying measures of agreement based on different measures of geographic distance).

WBS 8.105 - Matching Process Improvement (Cont.)

Near-term Project Focus (2):

- Determine high-level requirements and milestone schedule for Comparing Matching Systems subteam.
- Continue interaction with Geographic Support System Initiative to leverage their efforts for interactive/real-time address matching and geocoding.
- Continue interaction with the 2020 Improving Quality Control Team to determine what support is needed from the Matching Process Improvement Team for real-time matching of reinterview rosters.
- Review project risk mitigation plans and develop contingency plans.

WBS 8.107 - Administrative Records Modeling

Project Description:

- Consolidate and conduct research efforts for examining optimal scenarios of curtailing contacts during the Nonresponse Followup (NRFU) field operations with the supplemental use of administrative record sources.
- Project Objectives

1) Simulate how administrative sources can be used to supplement curtailed NRFU data collection.
   - Administrative records usage based on either a rule-based algorithm, model or both
   - Count and characteristic imputation approaches that can possibly include the use of administrative records sources

2) Assess the coverage and cost of simulated scenarios.

3) Compare resulting scenarios to identify possible combinations of NRFU visits and administrative records use to pursue.

WBS 8.107 - Administrative Records Modeling (Cont.)

Project Description (continued):

- Coordinate with the 8.102 Administrative Records Fitness-for-Use team based on their recommendations.
- Coordinate with the 5.101 Coding, Editing and Imputation team.
  - 8.107: research involving curtailed NRFU, imputation, and administrative records is focused on the next 6 months
  - 5.101: long-term focus on methods to improve count imputation and characteristic imputation, with and without the use of administrative records
- Coordinate with other teams researching field operation and budget implications.

WBS 8.107 - Administrative Records Modeling (Cont.)

Challenge Highlights:

- Can we have NRFU contact strategies where certain areas or housing units would receive very few or no field visits?
- Can we have NRFU contact strategies where different areas or housing units receive different number of visits?
- What are the coverage metrics necessary to assess simulations?

WBS 8.107 - Administrative Records Modeling (Cont.)

Recent Accomplishments/Near-Term Project Focus:

- Drafted and briefed the team's 2020 Research and Testing study plan.
- Shared initial research of curtailed NRFU contact strategies using both count imputation methods without administrative records and approaches that used administrative records.
- In April, report interim modeling results on ways to supplement curtailed NRFU strategies. Results to address:
  - + Implementing administrative record and non-administrative record sources for curtailed NRFU units (model-based vs. rule-based algorithms).
  - + Utilizing count and characteristic imputation methods to account for remaining census missing data.
  - + Determining metrics to assess the accuracy of a scenario of NRFU contact strategy and administrative record source usage.
  - + Estimating approximate cost savings for different NRFU contact strategies proposed.